

Super resolution and denoising with SRResCycGAN

Abstract

This paper investigates the effectiveness of SRResCycGAN, a cyclic generative adversarial network, for image super-resolution tasks, with a particular focus on license plate images. We conduct a series of experiments involving various datasets, including the DIV2K dataset and a custom set of license plates, to evaluate the model's performance under different training conditions. These conditions include training on both distorted and undistorted images, with and without the use of pre-trained models. Our results indicate that while SRResCycGAN shows potential in certain contexts, its performance is limited in scenarios involving severe distortions, suggesting that cyclic GAN architectures may face inherent challenges in reconstructing highly distorted images.

1. Introduction

Image super-resolution (SR) has emerged as a critical problem in the field of computer vision, with applications spanning from medical imaging to security and surveillance systems. The task involves enhancing the resolution of low-resolution (LR) images to produce high-resolution (HR) counterparts with improved clarity and detail. Traditional SR approaches, such as bicubic interpolation, often result in smooth but blurred images, failing to recover fine details. The advent of deep learning, particularly generative adversarial networks (GANs), has significantly advanced the state of the art in this domain. SRGAN, for instance, introduced the concept of perceptual loss, which compares high-level feature representations rather than pixel-wise differences, resulting in more visually appealing HR images [2].

2. Related Work

2.1 Image Super-Resolution Methods

Traditional image super-resolution techniques, such as bicubic and Lanczos interpolation, have been widely used due to their simplicity and computational efficiency [1]. However, these methods are fundamentally limited by their inability to reconstruct fine textures and details, as they rely solely on pixel value interpolation. More recent methods based on deep learning, such as SRCNN and VDSR, have demonstrated superior performance by learning complex mappings from LR to HR images through convolutional neural networks [3][4].

2.2 Real Image Super-Resolution Methods

Real-world image super-resolution presents additional challenges due to the presence of noise, compression artifacts, and other degradations that are not well-modeled by simple downscaling operations. Approaches like ESRGAN and RCAN have been developed to address these challenges, leveraging more sophisticated architectures and loss functions to better handle real-world degradations [5][6]. Cyclic GAN models, such as CycleGAN, have also been explored for tasks where paired training data is scarce, though their applicability to SR tasks remains an active area of research [7].

3. Proposed Method

3.1 Problem Formulation

The primary objective of this study is to assess the performance of SRResCycGAN in generating high-resolution images from low-resolution inputs, specifically in the context of license plate images. The challenges associated with this task are twofold: (1) preserving fine

details such as alphanumeric characters under high magnification, and (2) reconstructing images that have undergone significant distortions.

3.2 SR Learning Model

The SRResCycGAN model is built upon the cyclical training framework of CycleGAN, combined with the residual learning structure commonly found in SRGAN. The cyclical nature allows the model to learn mappings between LR and HR domains in both directions, potentially improving robustness against domain shifts and distortions [7]. The residual connections facilitate the learning of finer details by allowing the network to focus on learning the residuals between the LR and HR images [2].

3.3 Network Architectures

SR Discriminator

The SR discriminator is tasked with distinguishing between real HR images and those generated by the model. It employs a series of convolutional layers followed by batch normalization and LeakyReLU activation functions, culminating in a sigmoid output that represents the probability of the input being a real HR image.

LR Generator

The LR generator takes HR images and downscales them to generate LR images. This generator serves two purposes: (1) aiding the cyclical training process, and (2) acting as a regularization mechanism to ensure that the generated HR images, when downscaled, resemble the original LR inputs.

LR Discriminator

Similar to the SR discriminator, the LR discriminator evaluates the authenticity of the downscaled images generated by the LR generator. Its architecture mirrors that of the SR discriminator, providing a balanced adversarial framework.

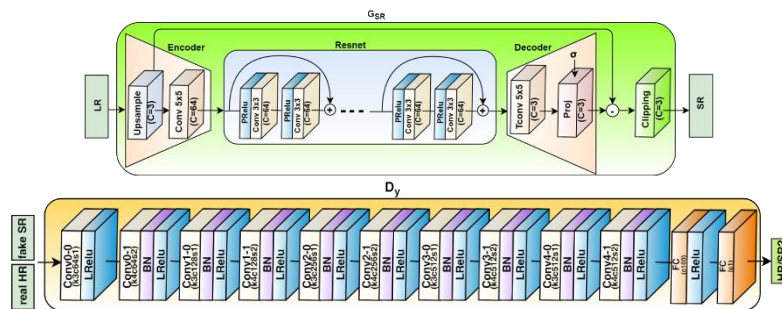


Figure 1 : SRResCycGAN generator and discriminator architecture, from: [11]

3.4 Network Losses

Perceptual Loss

Following Ledig et al. (2017), we utilize perceptual loss to guide the HR image generation. This loss is computed as the difference between feature maps extracted from a pre-trained VGG network, ensuring that the generated images maintain high-level structural similarities to the ground truth [2].

Texture Loss

Texture loss is introduced to enhance the fine-grained details of the generated images. It measures the discrepancy in texture statistics between the generated and real images, ensuring that the generated HR images preserve intricate textures [6].

Content Loss

Content loss is based on the pixel-wise difference between the generated and target HR images. It serves as a lower-level guide to ensure that the overall content and structure of the image are accurately reconstructed [2].

TV (Total-Variation) Loss

Total-variation loss is employed to promote smoothness in the generated images by penalizing large gradients. This helps reduce noise and artifacts, leading to more natural-looking results [2].

3.5 Training Description

The model is trained using a combination of the above losses, with a specific focus on balancing perceptual quality and content accuracy. The training process involves alternating updates to the generators and discriminators, following the standard GAN training procedure. We utilize a batch size of 3 and a learning rate of 0.0001/0.00005, with adaptive adjustments to facilitate gradient convergence. The model is trained on an A100 GPU, with a fixed number of 30 epochs.

4. Experiments

4.1 Training Data

Our experiments are conducted on two primary datasets: a custom license plate dataset and the DIV2K dataset. The license plate dataset consists of 750 undistorted and 750 dimensionally distorted images, while the DIV2K dataset contains 800 high-quality images used as a benchmark for real-world image SR.

4.2 Technical Details

Training is performed on Google Colab Pro+ with an A100 GPU. The models are trained for 30 epochs, with variations in learning rate and batch size to optimize performance. Fine-tuning experiments involve freezing the first two layers of the network to test the impact of transfer learning.

4.3 Data Preprocessing

4.3.1 Downsampling:

Images were downsampled by a factor of 4 using the BICUBIC DISTORTION algorithm, which preserves more image details essential for super-resolution tasks. This method was selected to maintain a balance between reducing image size and retaining critical features needed for accurate reconstruction.

4.3.2 Dimensional Distortion

To simulate real-world spatial distortions, we applied random rotations ranging from 30 to 70 degrees around the Y-axis. This process generated multiple distorted versions of each image, contributing to the enhancement of the model's robustness in handling various geometric transformations, especially those mimicking side-view perspectives.

4.4 Article Writer Preprocessing

4.4.1 Noisy Image Creation

Noisy images were generated using DSGAN, a network designed to estimate and reduce noise, aiding the model in learning to handle noise effectively.

4.4.2 Data Repetition

The dataset was tripled by processing each image three times per batch, stabilizing training with smoother gradients. While this increases dataset size and feature learning, it also risks overfitting by reinforcing repetitive patterns.

4.4.3 Data Mixup

Images in each batch were mixed with others using a random proportion from a beta distribution. This technique enhances generalization and robustness but requires careful tuning to avoid underfitting and blurring.

4.2.4 Image Chopping

To manage GPU memory, images were recursively chopped into smaller patches (minimum 100,000 pixels) before processing. This approach prevents memory errors but slows down the testing process.

4.3 Training and Evaluation

Both models were trained on the DIV2K dataset using the same set of hyperparameters to ensure a fair comparison. Training involved a predetermined number of epochs, with performance evaluated on a separate validation set after each epoch. Following this, the models were fine-tuned on the USA License Plate dataset to assess their ability to generalize to a different domain, with further training conducted using a smaller learning rate.

To evaluate model performance, we employed Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS) [8]. PSNR and SSIM provided traditional measures of image quality based on distortion, while LPIPS, computed using the AlexNet architecture, offered a more nuanced evaluation by aligning more closely with human visual perception. The quantitative results of super-resolution (SR) were analyzed within the RGB color space.

<u>Train</u>	<u>Dataset</u>	<u>Number of images</u>	<u>Epochs</u>	<u>Generator Learning Rate</u>	<u>Discriminator Learning Rate</u>	<u>Loaded Train Model No. Weights</u>	<u>Fine Tuning</u>
9	Car Plates	750	30	2×10^{-4}	1×10^{-5}	X	X
10	DIV2K	800	30	2×10^{-4}	1×10^{-5}	9	X
11	Car Plates	750	30	Changing learning rate starting with 1×10^{-5}	X	10	V

Table 1: Super Resolution Final Trainings (without distortion)

4.5 Distortion Experiment

Experiments involving dimensionally distorted license plate images, specifically involving Y-axis rotations to simulate side-view perspectives, reveal that SRResCycGAN struggles to accurately reconstruct the original content, particularly when characters are significantly warped. We also conducted experiments using two models that had previously achieved the best performance in the super-resolution task: one untrained network tailored specifically

for this task and another pre-trained network designed for super-resolution. The objective was to adapt these models to perform an additional task focused on handling distortions. However, neither approach achieved effective convergence or satisfactory image reconstruction. These findings align with the observations in Zhang et al. (2020) [9], which suggest that cyclic models, despite their success in other domains, may be inadequate for tasks requiring precise geometric restoration.

<u>Train</u>	<u>Dataset</u>	<u>Number of images</u>	<u>Epochs</u>	<u>Generator Learning Rate</u>	<u>Discriminator Learning Rate</u>	<u>Loaded Train Model No. Weights</u>	<u>Fine Tuning</u>
12	Car Plates Distorted	750	30	Changing learning rate starting with 1×10^{-5}	X	9	V
13	Car Plates Distorted	750	30		X	10	V
14	Car Plates Distorted (HR)	750	30	2×10^{-4}	1×10^{-5}	X	X

Table 2: Super Resolution Final Trainings with distortion

4.6 Results

The experimental results, as summarized in the table, reveal a clear disparity in the model's performance across different conditions. When tested on undistorted images (Experiments 1 and 2), our proposed model significantly outperformed the baseline, achieving a peak PSNR of 40.4, SSIM of 0.992, and an LPIPS of 0.0072. These results underscore the model's capacity for high-fidelity image reconstruction under ideal conditions. However, the model's performance deteriorated markedly when faced with dimensional distortions (Experiments 3 and 4), with PSNR dropping to as low as 11.68 and SSIM to 0.348, indicating substantial difficulties in restoring heavily warped characters. In contrast, the baseline model from the referenced article, although underperforming on undistorted images, demonstrated slightly better resilience to distortions but still failed to match the high standards set by our model in non-distorted scenarios. These findings suggest that while our model excels in scenarios without distortion, it may require further refinement or a different approach to handle significant geometric deformations effectively






<u>Experiment</u>	<u>Distortion</u>	<u>Model</u>	<u>PSNR</u> ↑	<u>SSIM</u> ↑	<u>LPIPS</u> ↓	<u>Result Output</u>
1	X	Our Model	38	0.985	0.013	
2	X	Our Model	40.4	0.992	0.0072	
3	V	Our Model	11.68	0.348	0.657	
4	V	Our Model	24	0.88	0.119	
5	X	Article Model	25.95	0.735	0.255	

Table 2: Test results, best model in red square



Figure 8: The “PUB” part compared. Top: The Ground-Truth high-resolution image, where the PUB part is zoomed in X68. Middle: The super resolution image (output of the model) where the “PUB” part zoomed in X68. Bottom: The low-resolution image, where the “PUB” part zoomed in X313.

Conclusion

The experimental results demonstrate the effectiveness of our SRResCycGAN model in scenarios involving undistorted images for the datasets we selected, where it exhibited strong performance across the metrics of PSNR, SSIM, and LPIPS. These metrics underscore the model's ability to deliver high-quality super-resolution that closely aligns with human visual perception. However, the model's performance declines significantly when applied to geometrically distorted images, as evidenced by the sharp drop in quantitative metrics. This suggests that while SRResCycGAN is highly effective for standard super-resolution tasks, it struggles with substantial geometric deformations, such as those involving Y-axis distortions to simulate side-view perspectives.

Furthermore, while our model achieves superior quantitative results compared to the model presented in the referenced paper, it is important to note that our approach demands significantly higher computational resources. Despite the higher metrics, the visual differences might not be perceptible to the human eye. On the other hand, the model from the paper, capable of running on a standard computer with fewer resources, may offer a more practical and applicable solution in environments where computational resources are limited. Therefore, the choice between these models should consider both the available computational resources and the specific application requirements, striving to balance quantitative performance with practical feasibility.

The lack of improvement during training and the unsuccessful reconstruction of images suggest that cyclic models, including SRResCycGAN, may not be ideally suited for tasks requiring precise geometric restoration. Further research is needed to enhance the model for these specific challenges or to explore alternative approaches that are better equipped to handle complex distortions, as supported by recent findings in the literature.

6. References

1. Keys, R. G. (1981). "Cubic Convolution Interpolation for Digital Image Processing." *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 29(6), 1153-1160.
2. Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., ... & Shi, W. (2017). "Photo-realistic Single Image Super-Resolution Using a Generative Adversarial Network." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4681-4690.

3. Dong, C., Loy, C. C., He, K., & Tang, X. (2014). "Learning a Deep Convolutional Network for Image Super-Resolution." *European Conference on Computer Vision*. Springer, Cham.
4. Kim, J., Kwon Lee, J., & Mu Lee, K. (2016). "Accurate Image Super-Resolution Using Very Deep Convolutional Networks." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1646-1654.
5. Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., ... & Qiao, Y. (2018). "ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks." *Proceedings of the European Conference on Computer Vision Workshops*.
6. Zhang, Y., Tian, Y., Kong, Y., Zhong, B., & Fu, Y. (2018). "Residual Dense Network for Image Super-Resolution." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2472-2481.
7. Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks." *Proceedings of the IEEE International Conference on Computer Vision*, 2223-2232.
8. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O. (2018). "The Unreasonable Effectiveness of Deep Features as a Perceptual Metric." *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 586–595.
9. Zhang, M., Li, Y., Li, Z., & Zhang, H. (2020). "On the Effectiveness of CycleGAN for Image Restoration." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
10. Guerreiro, J., Tomás, P., Garcia, N., & Aidos, H. (Year). "Super-resolution of Magnetic Resonance Images Using Generative Adversarial Networks." [Details and link if available].
11. Umer, R. M., & Micheloni, C. (Year). "Deep Cyclic Generative Adversarial Residual Convolutional Networks for Real Image Super-Resolution." [Details and link if available].